

EXHIBIT 4



Isaac Ray Forensic Group

65 E. Wacker Place, Suite 2240
Chicago, Illinois 60601
tel: 312.621.9002 fax: 312.621.9003

Shneur Nathan, J.D.
Natalie Adeeyo, J.D.
Nathan & Kamionski, LLP
33 W. Monroe, Ste. 1830
Chicago, IL 60603

June 08, 2023

Dear Mr. Nathan and Ms. Adeeyo:

At your request, I have reviewed the March 07, 2023 report offered by Plaintiffs' expert Richard Leo, Ph.D., J.D., as well as the police reports, documents, and various deposition and court transcripts provided to me, in the matters of John Fulton v. Robert Bartik et al., Case No. 1:20-cv-03118 and Anthony Mitchell v. Robert Bartik et al., Case No. 1:20-cv-03119. On August 31, 2006, John Fulton and Anthony Mitchell were convicted of the 3/10/03 murder and kidnapping of Christopher Collazo and sentenced to 31 years' incarceration. At the time of their arrests, Mr. Fulton was an 18-year-old student at De La Salle High School and Mr. Mitchell was 17 years old and not attending school. Since the time of their arrests, Mr. Fulton and Mr. Mitchell spent 16 years in custody before their post-conviction petitions were granted and the State dismissed their charges in 2019.

Following a reversal of their convictions, Mr. Fulton and Mr. Mitchell were released from the Illinois Department of Corrections and filed motions to obtain certificates of innocence. On 2/21/20, both were denied certificates of innocence, Judge Leroy K. Martin, Jr. finding that Mr. Fulton and Mr. Mitchell had not proven they were innocent of the offenses charged. As a result of the Court's finding, the falsity of Mr. Fulton's and Mr. Mitchell's confessions has not been definitively proven. In their pleadings, Mr. Fulton and Mr. Mitchell nonetheless allege that they were wrongfully convicted of killing Mr. Collazo because their confessions were false and coerced by the police.

The plaintiffs have retained Richard Leo, Ph.D. to offer research on false confessions and his opinions regarding the falsity of Mr. Fulton's and Mr. Mitchell's 2003 confessions. You have asked that I consider Dr. Leo's conclusions and opinions, including the methodology of his research into the phenomenon of false confessions (and that of the field of false confession research more generally), and the methodology he employed in arriving at opinions in Mr. Fulton's and Mr. Mitchell's cases specifically.

Below please find my conclusions and opinions in this matter, to a reasonable degree of certainty in the field of forensic psychology. Sources of information are included below, followed by an Extended Report section containing a detailed analysis of relevant investigations in the field.

Respectfully submitted,



Diana S. Goldstein, Ph.D., ABPP
Owner and Director of Neuropsychology
Licensed Clinical Psychologist, #071-006006
Isaac Ray Forensic Group, LLC

INFORMATION SOURCES

The following collateral documents were available for review:

1. Plaintiff John Fulton's Complaint
2. Plaintiff Anthony Mitchell's Complaint
3. Officer Robert Bartik's Polygraph Examination Reports (FULTON-MITCHELL 003199-003203)
4. Investigative File (CITY_FULTON_MITCHELL_000001-000118)
5. Defendant Officers' General Progress Reports (FULTON-MITCHELL 000013-000044)
6. CPD Cleared Closed Report (FULTON-MITCHELL 000001-000012)
7. Grand Jury Testimony of Johnnitta Griffin (3/14/2003)
8. Trial Testimony of Johnnitta Griffin (8/24/2006)
9. Trial Testimony of Detective Edward Winstead (8/25/2006)
10. Trial Testimony of Detective Edward Winstead (8/29/2006)
11. Trial Testimony of Detective James Breen (8/30/2006)
12. Trial Testimony of Detective Leonard Rolston (8/29/2006)
13. Trial Testimony of Detective Joseph Struck (8/29/2006)
14. Trial Testimony of Detective John Zalatoris (8/30/2006)
15. Motion to Suppress Testimony of Detective Girardi (2/28/2005)
16. Motion to Suppress Testimony of Detective Leonard Rolston (7/13/2005)
17. Motion to Suppress Testimony of Anthony Mitchell (7/13/2005)
18. Motion to Suppress Testimony of Officer Robert Bartik (4/28/2006)
19. Motion to Suppress Testimony of Detective Leonard Rolston (4/28/2006)
20. John Fulton's Verified Petition for Post-Conviction Relief (FULTON-MITCHELL 000647-000675)
21. Written Statement of Antonio Shaw (3/22/2003)
22. Deposition Testimony of John Fulton (9/19/2022)
23. Deposition Testimony of Anthony Mitchell (9/20/2022)
24. Deposition Testimony of Marcus Marinelli (12/17/2021)
25. Deposition Testimony of James Breen (8/3/2022)
26. Deposition Testimony of John Zalatoris (7/18/2022)
27. Deposition Testimony of Johnnitta Griffin (11/30/2022)
28. Deposition Testimony of Leonard Rolston (8/5/2022, 8/12/2022)
29. Deposition Testimony of Robert Bartik (6/22/2022)
30. Deposition Testimony of Antonio Shaw (10/21/2021)
31. Deposition Testimony of Ronald Smith (11/18/2021)
32. Deposition of Elliott Zinger (11/29/2022)
33. Videotaped Confession of Anthony Mitchell
34. Video from Mitchell Hospital Emergency Room (SAO 14688-14690)
35. Video from Lake Meadows Apartments (SAO 14688-14690)
36. Memorandum by Jake Rubinstein re John Fulton's Statement (SAO 002402-002407)
37. Official Statement of Facts by Nancy Nazarian (SAO 000550-000552)
38. Expert Report of Richard Leo, Ph.D., J.D., 3/07/23
39. Deposition of Richard Leo, Ph.D., J.D., Transcript and Exhibits, 5/03/23

OPINION AND COMMENTARY

Investigations into the phenomenon of false confessions among suspects in criminal matters are fraught with methodological problems. The empirical work remains in its nascent form, and because its data points are relatively hard to come by, the inferences drawn from this body of work must necessarily remain *qualitative* in nature, not quantitative; reliable conclusions cannot be drawn at this time. Some examples are provided here and discussed in greater detail in the Extended Report section, which follows.

1. Both field and experimental investigations into interrogation approaches suffer an irreparable flaw from a methodological standpoint—field research consisting of observing interrogators and surveying their practices lack the necessary controls to test causal relationships between interrogation approach and the reliability of the obtained information, whereas the reliability of obtained confessions is known in a controlled laboratory but lacks ecological validity. That is, the generalizability of an experiment with student volunteers to an actual interrogation in a criminal investigation is unknown but likely poor given the social, cognitive, and psychological processes underlying both interrogator and suspect behavior would reasonably be expected to differ in real contexts.
2. Field investigations (those of actual criminal suspects, convicted or exonerated individuals) are few in number and have failed to separate murder cases from other felony and misdemeanor cases, or failed to provide sufficient information on crime types at all. Indeed, no study has uniformly assessed murder suspects and their interrogations.
3. Field investigations of exonerees are based on a presumption of accuracy in innocence determinations (as well as falsity of confessions in some work). This is problematic as some exoneree samples (The Innocence Project) include cases involving pardons by Governors that have been rejected by the Courts, rendering the purity of these study samples inscrutable, if not suspect. The DNA-exonerated cases included in the National Registry of Exonerations offer a unique opportunity to study those who falsely confessed to murder but to date the numbers are too low for analysis and the endeavor has not been undertaken (1226 homicide cases as of the 2022 annual report, just 12% considered due to false confession).¹
4. Those whose false confessions are determined pretrial and do not progress to adjudication are rarely studied and are not compared to those whose cases do progress, thus field studies are largely based on a cross section of individuals whose cohort was determined by unknown/unquantifiable variables or moderators.²

¹ The National Registry of Exonerations 2022 Annual Report. Published May 8, 2023.

² In a 2022 Michigan News interview with University of Michigan Law Professor Samuel Gross discussing the 10-year mark of the National Registry of Exonerations, he was asked to describe the cases that get overturned. His response: “The most common causes of false conviction we know of from exonerations are perjury and other false accusations (61% of cases), governmental misconduct of many sorts (56%), mistaken identifications by eyewitnesses (27%), false or misleading forensic evidence (24%) and false confessions (12%)... These are the reasons for the false convictions we know about. There are many that we don’t know about, for one reason or another. I’d say that numerically, the most common cause of false convictions is the pretrial detention of defendants who can’t make bail. That confinement leads some innocent defendants in low-level cases to take plea bargains that will allow them to go home immediately, or in days, rather than wait for months in jail before trial. But that’s just an educated guess. We almost never see those cases among the exonerations we list. It’s very hard to obtain an exoneration after a guilty plea, and hardly anyone ever tries for a misdemeanor or low-level felony.” Reference: Exoneration Registry tracks 10 years of data, but innocent defendants won’t get day in court. Jared Wadley, April 12, 2022. Michigan News, University of Michigan.

5. Related, early definitions of false confession have been revised significantly (e.g., Dr. Leo developed but later abandoned his own system of characterizing false confessions), and he and others have repeatedly pointed out the unreliability of cases in their own study samples.
6. False confession researchers have drawn inferences from studies with small sample sizes. In Meissner et al.'s 2012 meta-analysis³ of false confession research (which included Dr. Leo's work), the authors bemoaned the small sample sizes in the work to date and called for researchers to make a concerted effort to do better:

“In our review of the available literature, a number of control variables could reasonably be included in such analyses (e.g., factors related to interrogator experience, crime type, interrogator/suspect ethnic backgrounds, geographic characteristics, etc.)... albeit many (if not all) of these studies may not have had sufficient sample sizes to consider multiple factors simultaneously. We strongly encourage researchers to obtain larger samples and initiate more systematic, multi-level analyses of the influence of interrogative methods.”
7. Researchers have similarly drawn conclusions from those that have not been replicated, or worse, *were* not replicated, including those related to age, sleep deprivation, ADHD and other “personal vulnerabilities” highlighted in the Fulton and Mitchell cases by Dr. Leo as risk factors for false confession. He refers to these factors as “generally accepted,” citing to the positive findings only. Experts in the field do not in fact universally agree on the nature of these factors, and many more have indicated they would not testify to them.⁴ As one example, more than 20% of experts did not agree that adolescents exhibit “immaturity of judgment,” which Dr. Leo describes in both Mr. Fulton's and Mr. Mitchell's cases, and 30% said they would not testify to it.
8. Related, researchers have overstated positive findings related to “personal vulnerability” factors and differences between interrogation techniques, that are either statistically insignificant (i.e., they show trends in the data only) and/or have small effect sizes, too small to be considered meaningful beyond a numbers' standpoint. Yet Dr. Leo applies such findings to the individual cases of Mr. Fulton and Mr. Mitchell, the foundation for which, in my opinion, is lacking.
9. As noted above, of the 1226 homicide cases listed in the National Registry of Exonerations as of the 2022 annual report, just 12% considered due to false confession. These data indicate that false confession cases comprise a minority of U.S. wrongful conviction samples and other factors are at play. Research on true confessions is scant, a factor that can't be overlooked; scores of interrogations are conducted daily among 18,000 law enforcement agencies across the United States. While we can't know the exact numbers because the rate of true confessions is not known, the likelihood that an interrogation will result in a false confession is relatively low, and that must be acknowledged, no matter how troubling the concept more generally. Yet Dr. Leo presumes false confession in Mr. Fulton and Mitchell's cases.
10. In his own work, Dr. Leo has moved from an intuitive definition of interrogation length (actual time in interrogation) to a nonintuitive one (total time in custody irrespective of interrogation duration). He describes the latter approach as generally accepted, yet I have seen no other researchers

³ Meissner C, Redlich A, Bhatt S, Brandon S. (2012). Interview and interrogation methods and their effects on true and false confessions. *Campbell Systematic Reviews*, DOI: 10.4073/csr.2012:13.

⁴ Kassir S, Redlich A, Alceste F, Lake T (2018). On the general acceptance of confessions research: Opinions of the scientific community. *American Psychologist*, 73(1), 63-80.

operationalize interrogation length in this confusing manner. More importantly, he provides no empirical basis for the conclusion they are one and the same.

11. Related, studies reporting on interrogation length have mixed misdemeanor and felony crimes and/or failed to isolate interrogation length in murder cases, almost certainly providing an underestimate of the average time spent in interrogation. In various reports on ranges of time actually spent in interrogation or estimated to have been spent by police interrogators (if reported at all), duration is upward of 4½ hours to 7 hours in some studies, yet no attempt has been made to identify and analyze those cases. Some researchers seem to count cases in which no real interrogation was required because the suspect confessed at the start of the interaction, skewing the data. Finally, the pool of reports from which these numbers are drawn is very small in any case, and some of them are not from the United States, where interrogation approaches differ. Dr. Leo relies on this small number of reports in concluding the interrogations of Mr. Fulton and Mr. Mitchell were comparatively “extraordinary” in length, yet the comparison is apples to oranges given his unique definition of interrogation duration was not employed in any of these studies.
12. Although Dr. Leo initially couches his opinions as being conditionally based on the truthfulness of Mr. Mitchell and Mr. Fulton, and states he is simply pointing out “risk factors” for false confession, he appears to lend unqualified credence to their accounts thereafter and goes on to draw seemingly causal lines from their purported false confessions to police interrogation tactics. In my opinion, support for such a relationship cannot reliably be established.
13. Finally, Dr. Leo’s personal analysis of the reliability of Mr. Fulton’s and Mr. Mitchell’s confessions (so-called “Indicia of Reliability,” p. 46 and p. 61) contain disputed facts between the parties. And while he claims not to be truth-seeking, his reliance on these disputed facts to support his opinion that Mr. Fulton and Mr. Mitchell’s confessions were indeed false meets the line, if not crosses it. The survey of confession experts referenced above indicates a lack of consensus on confessions being verifiable by details they contain about a crime (just 19% agreed this to be the case and 18% said they would testify to it in court). I venture no such opinion.

The study of human behavior has always been and will always be extremely challenging; any researcher, be they social scientist, psychologist, or any other empiricist trained in research design and statistics, knows unequivocally they are fortunate to account for just a small amount of its variance; too many variables comprise the complexity of it. Moving from the statistically significant to the clinically meaningful (or to any other applied setting, in this case the Courts) is a jump no conservative researcher should be willing to make without data that are plentiful, that have been subjected to the rigors of repeated replication for accuracy and error rate determinations, and across critical independent and moderator variables to ensure its generalizability. That cannot and does not exist here. I find no other way to offer my opinion but plainly: at the current time, false confession research has no place in a courtroom. Its potential to be misleading and confusing is too great.

EXTENDED REPORT

Below please find a critical analysis of a number of false confession investigations on which Dr. Leo relied to support opinions in the Fulton and Mitchell matters. Some additional investigations are included in order to illustrate the current state of the field and its challenges. For organizational purposes, the analysis is divided into subsections, each reflecting its own topic relevant to the false confession literature. An opinion related to Dr. Leo's work in the Fulton and Mitchell matters is included within each section.

1. Impact of Traditional versus Information-Gathering Interrogation Techniques on False Confessions

Opinion

Based on all of the considerations below, it is clear that at this time the empirical literature does not support the conclusion that traditional American interrogation techniques consistently lead to false confessions.

There are very few field studies that have attempted to do what Dr. Leo is endeavoring in the Fulton and Mitchell matters—to identify in their respective interview transcripts the “accusatory interview techniques”⁵ or other techniques he has personally identified and developed,⁶ in order to determine a seemingly causal association between a given technique, or set of techniques, and a false confession. Pearse et al. (1998) eloquently addressed the inherent problem in such an exercise while presenting findings from their London, England field study:

“It is important to realize that it is not possible to control for the ‘ground truth’ of the allegation in each case. Clearly, there are likely to be some suspects arrested by the police who will be innocent, and there will also be many who are guilty. The law in England and Wales, appreciating the inherent difficulty in attempting to achieve this elusive goal, has for many centuries sought to determine the more manageable concept of proving a case ‘beyond reasonable doubt’. This is not an option that is open to empirical research and such a limitation needs to be articulated.”

Because field studies lack “ground truth” (outcome knowledge of what *is* in fact a true versus false confession), and experimental studies have not reliably distinguished between them (see the 2019 experimental investigation by Salvati & Houck, below), the reliability of Dr. Leo's approach must be questioned.

In their 2012 meta-analysis of interview and interrogation techniques, Meissner et al.⁷ commented on the scant research into the method generally (p. 32):

“Our review of the available field literature located 33 potentially eligible observational studies on interrogation, though only 5 of these studies empirically assessed the relationship between interrogative approaches and elicitation of a confession.”

The authors further pointed out the absence of any reliable empirical investigation into the relationship between interrogation techniques and determining the likelihood of a *true versus false* confession among

⁵ A reference to the “Reid Technique” or “Reid Method”; Inbau F, Reid J, Buckley J, Jayne B. (2013). *Criminal interrogation and confessions* (5th ed.). Burlington, MA: Jones and Bartlett Learning.

⁶ Leo (1996). Inside the interrogation room. *Journal of Criminal Law & Criminology*, 86, 266-303.

⁷ Meissner C, Redlich A, Bhatt S, Brandon S. (2012). Interview and interrogation methods and their effects on true and false confessions. *Campbell Systematic Reviews*, DOI: 10.4073/csr.2012:13.

those five studies, among them Dr. Leo's study from 1996:⁸

"It is important here to note that field studies fail to offer us important information regarding the relative diagnostic value of the confession that is elicited. That is, such studies lack 'ground truth' that would enable us to factually determine the veracity of the statement provided by a suspect, and thereby preclude our ability to assess the diagnostic value of the information elicited and therein the effectiveness of such techniques when employed in the field."

Meissner noted that certain methods for assessing veracity have been employed in field studies, one being the strength of available evidence against a defendant (used by Behrman & Davey⁹ in eyewitness identification research and by Leo & Ofshe¹⁰ in false confession research—in Fulton and Mitchell's cases Dr. Leo analyzes the cases and points to "Indicia of Unreliability"), yet none of the investigations employed the method for study. The relationship between interrogation approach and likelihood of producing a credible confession, true or false, was simply untested at the time.

In 2018, Kassin et al. surveyed 87 confession experts, many of whom were highly published and had considerable courtroom experience, about their opinions on 30 concepts relevant to deception detection, police interrogations, confessions, and relevant general principles of psychology. They defined general consensus in the field as 80% agreement or higher. While 85% of experts agreed an information-gathering versus confrontational approach to interrogation produced more diagnostic outcomes (i.e., separated true from false confessions at a higher rate), just 76% said they would testify to this in court.

In 2019, Salvati and Houck¹¹ investigated the relationship between interrogation techniques and false confessions using an experimental paradigm. Participants (150) recruited from Amazon's Mechanical Turk system were randomly assigned to complete one of two simulated interrogation exercises designed to model different interrogation tactics: a) the "accusatorial" Reid Technique; or b) the information-gathering "Compliance" model. Participants read about a crime they were hypothetically accused of committing and were instructed to maintain their innocence while responding to an interrogator. False confession rates were examined. At the end of the interrogation, all subjects were prompted to confess, at which time they provided open-ended written confessions and wrote about their attitudes toward their confession and the interrogator. The emotional and cognitive themes in these written confessions and open-ended responses were examined. Results revealed no differences in false confession rates between conditions. However, participants exposed to Reid-based questioning used relatively more negative, angry, and "close-minded (cognitively simple)" language when describing their interrogator and less certainty in confessions.

In 2021, Houck et al.¹² noted the field's general hypothesis and sentiment that non-coercive, rapport-based strategies are favorable for effective interrogation whereas psychological coercion reduces the likelihood of obtaining reliable information, suspect cooperation and true versus false confessions,

⁸ Leo (1996). Inside the interrogation room. *Journal of Criminal Law & Criminology*, 86, 266-303.

⁹ Behrman B, Davey S. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law & Human Behavior*, 25, 475-491.

¹⁰ Leo R, Ofshe R. (1998). The consequences of false confessions: Deprivations of liberty and miscarriages of justice and in the age of psychological interrogation. *Journal of Criminal Law & Criminology*, 88, 429-496.

¹¹ Salvati J, Houck S (2019). Examining the Causes and Consequences of Confession-Eliciting Tactics during Interrogation. *Journal of Applied Security Research*, 14(3), 241-256.

¹² Houck S, Salvati J, Vrij A, Newman L. (2021). Simulating interrogation in the lab: Examining the effectiveness of physical pain, coercive verbal pressure, and rapport-building for obtaining reliable information. *Journal of Applied Security Research*, DOI: 10.1080/19361610.2021.1949954.

among other things. However, the authors also reflected on the challenges of drawing reliable inferences from this research:

“However, both the field and laboratory research from which these conclusions are derived are susceptible to reasonable criticisms. Field research draws from real-world interrogation contexts: interviews with interrogators and surveys of their practices (Cleary & Warner, 2016);¹³ evaluations of interrogations, including those conducted in false confession/wrongful conviction cases (Drizin & Leo, 2004);¹⁴ and expert accounts of terrorist interrogations (Alison et al., 2014).¹⁵ Although this research is highly ecologically valid, it lacks the necessary experimental controls to test causal connections between coercive interrogation and key outcomes (e.g., the reliability of intel). To compensate for this limitation researchers have developed innovative experimental laboratory procedures to model certain forms of interrogation under highly controlled conditions... The drawback of these approaches is their limited ecological validity. That is, the social, cognitive, and psychological processes underlying interrogator and suspect behavior may operate differently in the laboratory than in real contexts.”

The authors investigated interrogation strategies that foster versus hinder obtaining reliable information. Comparing physical pain (submerging one’s hand in very cold water to emulate torture), psychological manipulation (accusatorial interrogation/coercive verbal pressure) and information-gathering (rapport-building in interrogation) in an experiment in which participants were interrogated about the location of hidden money, they found a significant, albeit small, effect across conditions ($\chi^2 \phi^2 = .29, p = .001$).¹⁶ Post hoc analysis revealed that participants offered unreliable responses significantly more often under coercive pressure conditions than under rapport-building conditions (41.8% versus 10.9%). However, the effect size was quite small (Odds Ratio = .17),¹⁷ and is therefore not particularly meaningful. In addition to pointing out that the rapport-building condition led to 11% of subjects providing unreliable information, the authors also noted that no confessions were obtained (instances where participants voluntarily gave up the location). They concluded that “although rapport is typically a safeguard for obtaining reliable intel, practitioners should understand it can lead to some inaccuracies.”

II. The Unreliability of Investigation Samples Opinion

A review of the false confession literature reveals a significant “impurity of sample” problem, making inferences drawn from such investigations unreliable.

1. Leo R, Ofshe R (1998). Consequences of False Confessions: Deprivations of Liberty and Miscarriages of Justice in the Age of Psychological Interrogation. *Journal of Criminal Law and Criminology*, 88(2), 429-496.

¹³ Cleary H, Warner T. (2016). Police training in interviewing and interrogation methods: A comparison of techniques used with adult and juvenile suspects. *Law and Human Behavior*, 40(3), 270. DOI: 10.1037/lhb0000175.

¹⁴ Drizin, S. A., & Leo, R. A. (2004). The problem of false confessions in the post-DNA world. *North Carolina Law Review*, 82(3), 891-1007.

¹⁵ Alison L, Alison E, Noone G, Elntib S, Waring S, Christiansen P. (2014). The efficacy of rapport-based techniques for minimizing counter-interrogation tactics amongst a field sample of terrorists. *Psychology, Public Policy, and Law*, 20(4), 421. DOI:10.1037/law0000021.

¹⁶ A phi value of .1 is considered a small effect,.3 a medium effect, and .5 a large effect.

¹⁷ An odds ratio is used to determine if a particular variable is a risk factor for the outcome of interest and provides the magnitude of its risk. Here, the Odds Ratio compares the relative odds of providing reliable vs. unreliable information given exposure to the variable of interest (e.g., interrogation condition). An Odds Ratio of 2 is the recommended minimum effect size representing a “practically” significant effect for social science data, 3.0 is a moderate effect, and 4.0 is a strong effect.

- A. Leo and Ofshe describe the manner in which they composed their original sample of 60 exonerees for study, which is problematic (p. 435-436):

“Cases of disputed confessions were identified through multiple sources: electronic media database searches; directly from case files; and from secondary sources...Based on the information that we obtained and reviewed, all of the cases studied satisfy the following conditions: no physical or other significant and credible evidence indicated the suspect’s guilt; the state’s evidence consisted of little or nothing more than the suspect’s statement “I did it;” and the suspect’s factual innocence was supported by a variable amount of evidence—often substantial and compelling—including exculpatory evidence from the suspect’s post admission narrative. For every case included in this study, there was no credible evidence corroborating the defendant’s “I did it” admission or supporting the conclusion that he was guilty.”

In the footnotes of pages 435-436, the authors provide further information about their data-gathering and methodological approach to verifying the “factual innocence” of their cases. This more detailed explanation illustrates presumption formed the basis for inclusion in the sample in an unknown number of cases; that in some cases the authors relied on sources of information from interested parties involved in ongoing litigation—some on which they served as paid consultants—that had the potential to produce biased or partial information regarding “innocence”; that the authors applied their own untestable litmus test of reliability to certain case evidence that challenged their determination of an individual’s innocence as “fact”; and that the authors came to have knowledge of unreliable evidence of “innocence” in certain cases yet maintained them in the sample for study:

“Due to the difficulty in obtaining case materials—especially in lesser known cases—all social science and legal research on miscarriage of justices [sic] relies on both primary and secondary source materials...The research reported here is no different. By necessity, we rely on a variety of sources to document our assertions of fact. Where possible, we have tried to draw directly on interviews, police transcripts, and trial records, but in many instances we were only able to obtain newspaper and magazine accounts, appellate court opinions, academic journal articles, and/or books...

The authors obtained case file materials (either substantial or selected portions) directly from the attorney(s) representing the confessor in 17 cases...The confessors’ attorneys typically requested consultation at a suppression hearing and/or criminal trial, during the post-conviction appeal, or in a civil proceeding following the termination of criminal charges. In one case, a governor requested consultation in connection with a pardon under consideration...

In many of the cases identified in this paper, the suspect supposedly also confessed to so-called “jailhouse snitches”—at the same time that he was busy recanting his uncorroborated confession to everyone else. Because jailhouse snitches stand to gain material concessions and sentence reductions, we do not regard their testimony as credible...

The amount of information on these cases varies. The analysis of some cases was based on access to virtually the entire case file, while the analysis of other cases was limited to journalists’ accounts or published appellate court opinions. Based on the available sources, no credible evidence supporting the confessor’s guilt was discovered in any of the cases reported in this article. Some investigations, however, involved questionable evidence that later proved to be unreliable.”

Finally, the authors convey their own sense of unreliability about the sample in their breakdown of cases, classifying 34 of the 60 cases (57%) as “proven false confessions,” 18 (30%) as “highly probable false confessions,” and the remaining 8 (13%) as “probable false confessions,” and of possible contamination within the sample, comment (p. 437):

“We recognize that for any case that could not be classified as a proven false confession, there is a possibility that our classification of the case might be in error. Despite strong evidence supporting the conclusion that the confession is false, it remains theoretically possible that one or more of the defendants we classify as false confessors may have committed the crime. Nevertheless, we believe that the disputed confessions discussed in this article would be judged false by an overwhelmingly [sic] majority of neutral observers with access to the evidence we reviewed.”

- B. Leo and Ofshe admit that the case data collected do not lend themselves to psychometric analysis (p. 435-436):

“The sixty cases discussed below do not constitute a statistically adequate sample of false confession cases. Rather they were selected because they share a single characteristic: an individual was arrested primarily because police obtained an inculpatory statement that later turned out to be a proven, or highly likely, false confession.”

Yet the authors arrive at conclusions from the data with great certainty, with little in the way of conservative interpretation or customary cautions about preliminary data and the need for replication. The reason a sample size of 60 does “not constitute a statistically adequate sample of false confession cases” is that it is simply too small to analyze statistically for any reliable (and therefore replicable) conclusions, even before its breakdown into three very small subgroups. Even if one assumes every case in the sample is in fact an innocent, and that the reason for wrongful conviction is a false confession (as opposed to any number of other contributing factors that resulted in reverse conviction), there is no control group to which to compare sample characteristics of interest to the authors, say the frequency with which the “risk factors” for false confessions they developed from this sample—which have formed the foundation for their opinions ever since—are found among “True Confessors”.

- C. Leo and Ofshe have determined no base rates of their identified “risk factors” among true confessors, thus the norm for true confessions among those with mental retardation, lower IQ or low cognition, mental illness, younger age, etc., remains unknown, and therefore the testability of the conclusion that such factors create significant relative risk for false confession (from a statistical standpoint) remains impossible. Clearly what would be required to test the reliability of Leo and Ofshe’s identified false confession risk factors is a research undertaking of considerable proportion. Nonetheless, this is what is required. This is particularly the case given the authors assign high levels of confidence to their research findings, offer them as evidence in applied settings such as the courts, and employ burden of proof language that has great potential to be misleading to fact-finders (p. 437):

“For the eighteen cases classified as *highly probable* false confessions, the evidence overwhelmingly indicated that the defendant’s confession statement was false... Thus, the defendant’s statement is classified as a highly probable false confession because the evidence led to the conclusion that his innocence was established *beyond a reasonable doubt*.”

Similarly:

“For the eight cases classified as *probable* false confessions, no physical or other significant credible evidence supported the conclusion that the defendant was guilty... Cases are included in this category if the *preponderance of the evidence* indicated that the person who confessed was innocent.”

2. Drizin S, Leo R (2004). The Problem of False Confessions in the Post-DNA World. *North Carolina Law Review*, 82(3), 891-1008.
- A. Drizin and Leo describe the manner in which they composed their second sample of exonerees for study, and the ways it differed from the 1998 sample. While larger (125 versus 60 cases), independent from the original set of cases, and purportedly containing “only interrogation-induced false confessions that can be classified as ‘proven’—that is, confessions that are indisputably false because at least one piece of dispositive evidence objectively establishes, beyond any doubt, that the confessor could not possibly have been the perpetrator of the crime,”¹⁸ the sample nonetheless remains problematic, which the authors acknowledge to some extent (p. 924):

“To assemble this cohort, the authors systematically identified disputed confessions primarily through electronic media and legal database searches and secondarily from other sources of information—such as police reports, trial transcripts, articles, and books discovered directly or brought to the authors’ attention by others. Once the authors identified a disputed confession case, we attempted to compile as much source material on the case as possible. In all cases, the authors made an attempt to locate and then contact the confessor’s attorney and, in some cases, the courthouse in which the case went to trial (if, in fact, it did go to trial). In some cases, we were able to compile vast amounts of information on the disputed confessions, such as interrogation transcripts and (audio or video) tapes, police reports, preliminary hearing and other pre-trial hearing transcripts, interviews with police interrogators and suspects, trial transcripts, depositions, published appellate court decisions, and other case materials. In other cases, we were left only with the facts reported in newspaper stories, despite our determined efforts to acquire more information about the case.”

As in the 1998 article, the footnotes contain more troubling information about the authors’ methodological approach to verifying innocence in their cases; the authors’ claim of purity in their “proven” false confession sample was unfortunately compromised by the willingness to retain unverifiable cases (p. 924):

“We have made every effort to independently verify and cite check the facts in the 125 proven false confessions reported in this paper. Regretfully, however, there were many cases—particularly ones in which police and prosecutors acknowledged that the confession was false and thus no trial occurred—for which we were unable to locate any other materials. In addition, many attorneys did not return phone calls seeking additional information on the false confession cases that we discovered. In those cases in which our phone calls were returned, some attorneys either had removed their files or were not able to locate them without great difficulty or expense. Even where old case files were accessible to the attorneys, the cost of reproducing case files was often prohibitive.”

¹⁸ The four criteria for a “Proven False Confession” are (Leo report, p. 6-7): 1) When it can be objectively established that the suspect confessed to a crime that did not happen; 2) When it can be objectively established that it would have been physically impossible for the confessor to have committed the crime; 3) When the true perpetrator is identified and his guilt is objectively established; and/or 4) When scientific evidence dispositively establishes the confessor’s innocence.”

The authors clearly appreciate the challenge to obtaining an uncontaminated sample yet make no attempt to limit or qualify the conclusions they (or others) may draw from it (p. 926-927):

“The authors’ decision to include only ‘proven’ false confessions in this Article introduces a conservative bias into our sample, for only a small number of cases involving a disputed confession come with independent case evidence that allows the suspect to prove his or her innocence beyond dispute. Most do not. Actual innocence, as commentators have repeatedly pointed out, is very difficult to verify. It is the rare disputed confession that can be proven indisputably false, because attempting to do so is akin to the proverbially difficult task of proving the negative.”

3. Gross S, Jacoby K, Matheson D, Montgomery N, Patel S (2005). Exonerations in the United States, 1989 through 2003. *Journal of Criminal Law and Criminology*, 95, 523-553.

Similar to Leo and Ofshe’s samples Gross et al.’s (2005) description of their sample of 340 exonerees between 1989-2003 (cases that are now a part of The National Registry of Exonerations) reflects impurity (p. 524):

“As we use the term, ‘exoneration’ is an official act declaring a defendant not guilty of a crime for which he or she had previously been convicted. The exonérations we have studied occurred in four ways: (1) In forty-two cases governors (or other appropriate executive officers) issued pardons based on evidence of the defendants’ innocence. (2) In 263 cases criminal charges were dismissed by courts after new evidence of innocence emerged, such as DNA.¹⁹ (3) In thirty-one cases the defendants were acquitted at a retrial on the basis of evidence that they had no role in the crimes for which they were originally convicted. (4) In four cases, states posthumously acknowledged the innocence of defendants who had already died in prison.”

To their credit, Gross et al. acknowledge problems with the composition of their sample (p. 525):

“There is no national registry of exonérations, or any simple way to tell from official records which dismissals, pardons, etc., are based on innocence. As a result, we learned about many of the cases in our database from media reports” (p. 525)...

Needless to say, we are in no position to reach an independent judgment on the factual innocence of each defendant in our data. That is not our purpose in this report. Instead, we look at overall patterns in the exonérations that have accumulated in the past fifteen years and hope to learn something about the causes of false convictions, and about the operation of our criminal justice system in general. It is possible that a few of the hundreds of exonerated defendants we have studied were involved in the crimes for which they were convicted, despite our efforts to exclude such cases” (p. 526-527).

Gross et al. also acknowledge that their initial data set had reliability issues (p. 525):

“An earlier version of this paper was released in April 2004, listing a total of 328 exonérations. See Samuel R. Gross, Exonerations in the United States, 1989 Through 2003 (April 9, 2004) (early unpublished manuscript, at <http://www.law.umich.edu/newsandinfo/exonerations-in-us.pdf>). After that report was released, we learned of about fifteen additional exonérations between 1989 and 2003, mostly by way of e-mails from individuals who contacted us about

¹⁹ Review of Gross et al.’s sample reveals 144 (42%) of the 340 cases were exonerated by DNA.

cases we had missed. We have also excluded three cases we listed in the initial report because additional information revealed that the defendants had not been ‘exonerated’ as we define the term.”

Finally, Gross et al. offer no explanation as to manner of diagnosis, but report 16 (5%) of the 340 cases in the sample to be mentally retarded, and another 10 (3%) “appear to have been suffering from mental illnesses.” The authors go on to describe false confession rates in the remaining cases, but in so doing describe uncertainty about the integrity of these analyses and their sample in general p. 545):

“Among all other exonerees (some of who [sic] may also have suffered from mental disabilities of which we are unaware) the false confession rate was 11% (33/313²⁰).”

4. Dr. Leo’s Deposition, *Jesús Sanchez v. The Village of Wheeling, et al.*, Case No. 19-cv-2437, 1/26/22.

Dr. Leo’s recent testimony in this matter, in which this author was involved, echoes the sample composition problems described above (p. 81-82):

Q. And how did you go about ascertaining that in these 125 cases the four criteria were met?

A. Well, we gathered as many case materials as we could on these cases. The case materials were the police reports; interrogation tapes and transcripts, where they existed; pretrial and trial transcripts; media reports or investigative reports; materials collected by The Innocence Project, or innocence projects; or lawyers. So we -- we tried to gather as many case materials as possible, and then we reviewed and analyzed those case materials. And those cases that met the definition, one of the -- one or more of the four prongs, we would have included; and those cases that did not, we would have excluded.

Further (p. 84-85):

Q. And when deciding whether or not any of the cases fit the criteria, one of the four criteria that you previously described, did you rely solely on undisputed facts?

A. I -- I don’t recall. There may have been disputed facts in some of these cases.

Further (p. 95):

Q. As we sit here today, are you able to tell us -- if we were to go through the list of the 125 individuals identified in this article, would you be able to tell us which criteria was met that allowed you and Mr. Drizin to conclude that the cases fit within the definition of a proven false confession?”

A. For some of the cases that I know better than other cases, I could. For example, the Norfolk Four case that I wrote a book about, co-wrote a book, I have good working memory of -- of those case facts. The Central Park jogger case, a famous case involving five African American juvenile defendants, I see that one of those was listed in 118. And then some of the other cases that I know better through my research or because I was involved at some point in the case as a

²⁰ There would appear to be 314 cases, but the percentage remains unchanged, both rounding to 11%.

consultant or expert witness, or if I taught the case or wrote about the case, many of the cases I would be able to do that for. Many of them I would not.

Additional testimony reveals that the reliability of Leo and Ofshe's methodologies can no longer be independently tested, nor the data replicated, as the authors did not maintain their data (p. 91-92):

Q. What happened to the case materials that you and Mr. Drizin relied on to conclude that each of these 125 cases fit the criteria for proven false confessions?

A. Well, I can only speak for myself, but I know that from time to time I need to dispose of case materials. As you can see, I have many file cabinets behind me. And I just collect too many documents. And this was a study published now 18 years ago. So I would have just destroyed materials that I no longer needed if I no longer had the storage capacity for them. Back then we didn't electronically store materials the way we do today, so everything was by paper.

III. The Lack of Specific Information about Homicide Suspects from Field and Experimental Studies Opinion

Dr. Leo relies on field and experimental investigations to support conclusions about Mr. Fulton and Mr. Mitchell's interrogations, vulnerabilities, and behavior in 2003. Yet these investigations tell us little about murder suspects and their interrogations as they have not been investigated as a specific group and most field samples contain few, if any, murder suspects. It is my opinion that it is premature to generalize findings from the available empirical literature to the Fulton and Mitchell cases.

1. In his first published field study,²¹ Dr. Leo observed and coded 122 live interrogations and 60 videotaped interrogations across three separate police stations in Northern California. He noted at the time, "There exist no contemporary descriptive or analytical studies of routine police interrogation practices in America." Dr. Leo described the typical felony suspect in his sample as a "young, lower or working class, African-American male," though the similarities to Mr. Fulton and Mr. Mitchell end there. Only 12.09% (22 subjects) of the sample involved murder cases, and 58% had prior felony records, which Mr. Fulton and Mitchell did not.
2. In his second published field study, Drizin and Leo²² studied 125 individuals they identified and believed had made "proven false confessions". The problematic manner in which the study sample was assembled is noted in the prior section above, *The Unreliability of Investigation Samples*. This sample contained 101 murder cases, comprising 81% of the sample. This was an excellent opportunity to assess characteristics of murderers as a group. However, the authors did not do so, or could not do so, likely because they were missing so much case data. For example, information about interrogation length, time in custody, and post-conviction outcome (e.g., whether the person remained incarcerated, had died in prison or was released) was available in just 44 cases, and reasons for release were available in only 37 cases. What the authors did conclude, unfortunately, could not in fact be determined from their study, as the sample was not randomly assembled (p. 946):

"As can be seen, the overwhelming majority of false confessions 81% occur in murder cases. The second largest category is rape (9%), followed by arson (3%). Not surprisingly, false confessions tend to be concentrated in the most serious and high profile cases, lending credence

²¹ Leo, R (1996). Inside the interrogation room. *Journal of Criminal Law and Criminology*, 86, 266-303.

²² Drizin, S. A., & Leo, R. A. (2004). The problem of false confessions in the post-DNA world. *North Carolina Law Review*, 82(3), 891-1007.

to the argument that false confessions—as well as wrongful convictions based on false confession—are more likely to occur in the most serious cases because there is more pressure on police to solve such cases.”

3. As noted in the Opinions section, in Meissner et al.’s 2012 meta-analysis²³ of false confession research (which included Dr. Leo’s work), in addition to the authors bemoaning the small sample sizes in the work to date, they pointed out that many variables, including crime type, could have been identified and studied, but had not been:

“In our review of the available literature, a number of control variables could reasonably be included in such analyses (e.g., factors related to interrogator experience, crime type, interrogator/suspect ethnic backgrounds, geographic characteristics, etc.)...”

4. Also noted in the Opinions section, the DNA-exonerated cases included in the National Registry of Exonerations offer a unique opportunity to study those who falsely confessed to murder but to date the numbers are too low for the multivariate analysis necessary for reliable results and the endeavor has not been undertaken (1226 homicide cases as of the 2022 annual report, just 12% considered due to false confession, or 147 cases).

IV. Lengthy Interrogations as a Risk Factor

In describing “lengthy (incommunicado) interrogation” as a situational risk factor that could “overbear a suspect’s will” and lead to a false confession, Dr. Leo criticizes the Detectives in Mr. Fulton’s and Mr. Mitchell’s cases and states Mr. Fulton’s “100-hour period of detention/custody and interrogation is considered extraordinarily lengthy” (p. 37). He explains on p. 36:

“Researchers consider the length of an interrogation to include both the time that a suspect is being questioned and/or accused as well as any breaks between questioning/accusation sessions, because breaks between accusation and questioning add to the stress and fatigue of the interrogation and sometimes are used as an interrogation technique itself. It is the total amount of time in custody during interrogation that matters.”

He goes on to note (p. 37):

“Empirical studies indicate that the overwhelming majority of routine custodial interrogations last less than one hour, whereas the combined time period of custody and interrogation in most interrogations leading to a false confession is more than six hours.”

Several points are relevant here:

1. In his 2004 paper (Drizin & Leo), citing his early 1996 work, he contradicts himself (p. 948):
 “...compared to studies of routing police interrogations in America, which suggest that more than 90% of normal interrogations last less than two hours.”
2. More importantly, Dr. Leo provides no citation for a definition of interrogation as including both actual interrogation time and other time in custody or at a police station. It was in his 2004 study that he seemingly quantified interrogation length in this manner for the first time (he provides little detail in that paper but describes interrogations lasting between 72-96 hours, indicating this must

²³ Meissner C, Redlich A, Bhatt S, Brandon S. (2012). Interview and interrogation methods and their effects on true and false confessions. *Campbell Systematic Reviews*, DOI: 10.4073/csr.2012:13.

have been total time in custody), without explanation. His description suggests such a definition is a generally accepted one, yet it defies logic, is not employed routinely in other investigations (this author has only seen the definition in Dr. Leo's work), and Dr. Leo did not use this definition in his original 1996 work.

3. Regarding the contention that "the overwhelming majority of routine custodial interrogations last less than one hour," there are several issues.
 - A. "Routine custodial interrogations" is not defined, but one can reasonably assume this does not include interrogations in murder cases, which is the only relevant type of interrogation here. We should therefore look to studies that have quantified interrogation duration solely in murder cases in order to assess whether Mr. Fulton or Mr. Mitchell's interrogation times are, in fact, extraordinary. Unfortunately, Dr. Leo offers no empirical evidence to support the conclusion. In his 1996 study, he reported the range of interrogation times across 182 felony cases, without breaking down case types, and just 22 in the sample were murder cases. His own data reflect significant variability, suggesting a breakdown by case type may have been relevant, yet he did not analyze the data (p. 279):

"Of course, the interrogations in my sample also varied by length, ranging from literally seconds (when the suspect invoked before the detective even introduced himself) to four and one-half hours... more than 70% of the interrogations in my sample lasted less than an hour, and only 8% lasted more than two hours."

Who are the approximate 15 individuals whose interrogations lasted more than two hours, and what was their average interrogation time as a group? Is it a fair representation of the data to suggest that most interrogations like the ones Fulton and Mitchell underwent should be expected to last under an hour?

- B. Other investigations commonly cited in support of brief interrogations being the norm suffer the same problem, and others. In Drizin & Leo's 2004 study, the authors admit (p. 948):

"Unfortunately, the length of interrogation either was not reported or could not be determined in almost two-thirds (65%) of the sample."

This left a small sample of 44 subjects. They go on:

"Of the cases in which the length of interrogation was either reported or could be determined, 16% lasted less than six hours; 34% between six and twelve hours; 39% between twelve and twenty-four hours; 7% between twenty-four to forty-eight hours; 2% between forty-eight and seventy-two hours; and 2% between seventy-two and ninety-six hours. These numbers are staggering. More than 80% of the false confessors were interrogated for more than six hours, and 50% of the false confessors were interrogated for more than twelve hours. The average length of interrogation was 16.3 hours, and the median length of interrogation was twelve hours. These figures are especially striking when they are compared to studies of routine police interrogations in America, which suggest that more than 90% of normal interrogations last less than two hours."

The comparison here is apples to oranges; Dr. Leo's 1996 sample consisted of interrogations he personally observed, and the durations he reported did not additionally include time in custody. It is unclear how the definition of "interrogation length" changed from actual time being questioned (as was used in Leo's 1996 study) to total time spent in custody. The only reference to the issue is contained in a footnote and suggests the change was one of convenience (p.849):

“We were only able to obtain the length of reported interrogation in slightly more than one-third of the cases we studied. In some of these cases, we possessed case records from which we could deduce the length of actual interrogation. In other cases, however, we were only able to obtain newspaper accounts of the reported length of the interrogation and were not able to independently verify their accuracy. One potential problem that concerned us is that lawyers will sometimes report the length of time in custody as the length of actual interrogation. As a result, the length of interrogation may be disputed between the parties... If police electronically recorded interrogations from start to finish, there would be no disputes about how long the suspect was questioned.”

Here the authors admit to mixing definitions in their sample (of 44 subjects for whom interrogation length could be assessed), some cases using actual time questioned, others total time in custody. The problem here is an obvious one, and no reliable data interpretation can be made. It is interesting to note that forward of this study, Drizin and Leo adopted and refer to the definition of interrogation duration as *total time in custody*. Such a definition was never empirically vetted against *actual time in interrogation* and cannot be presumed to measure the same thing.

- C. This misrepresentation of what is known (or generally accepted) regarding interrogation length is perpetuated in the “White Paper” published by Leo and his colleagues in 2010²⁴ (p. 16):

“Observational studies in the U.S. and Britain have consistently shown that the vast majority of interrogations last approximately from 30 minutes up to 2 hours (Baldwin, 1993; Irving, 1980; Leo, 1996b; Wald et al., 1967). In a recent self-report survey, 631 North American police investigators estimated from their experience that the mean length of a typical interrogation is 1.60 hours... these same respondents estimated on average that their longest interrogations lasted 4.21 hours (Kassin et al., 2007)... In their study of 125 proven false confessions, Drizin and Leo (2004) thus found, in cases in which interrogation time was recorded, that 34% lasted 6–12 hours, that 39% lasted 12–24 hours, and that the mean was 16.3 hours.”

Again, Drizin and Leo’s 2004 sample cannot be compared to any of these investigations as an unknown number of their cases also included time in custody when no interrogation was taking place, a definition not used by other researchers.

- D. A review of several of the references from Kassin et al.’s 2010 “white paper” reveals significant problems with relying on the empirical work of Baldwin (1993), Wald et al. (1993) and Kassin et al. (2007) to support the conclusion that in the U.S. and Britain “the vast majority of interrogations that from approximately 30 minutes to up to 2 hours.” This interpretation of the articles is misleading.
- i. Baldwin’s (1993)²⁵ investigation involved reviewing 600 audio and video tapes recorded in three police forces in 1989 and 1990 to examine the way in which interviews with suspects were conducted. A look at the study sample reveals one that cannot be compared to murder suspects undergoing interrogation. Nor therefore can the length of interrogations be compared to Mr. Fulton’s or Mr. Michell’s interrogation. Baldwin notes:

“The sample of 400 video recordings that was used in this enquiry did not constitute a representative cross-section of all interviews since its composition was influenced by whether or

²⁴ Kassin S, Drizin S, Grisso T, Gudjonsson G, Leo R, Redlich A (2010). Police-induced confessions, risk factors, and recommendations. *Law and Human Behavior*, 34, 3-48

²⁵ Baldwin, J. (1993). Police interview techniques: Establishing truth or proof? *British Journal of Criminology*, 33, 325–352.

not officers and suspects were prepared to have the interview video recorded... There were also variations from station to station in the types of case that were recorded. The 400 cases make up, then, something of a hotch-potch, but the samples do include matters of all levels of seriousness and hundreds of different officers. A few grave offences were included in the sample, but the great mass were relatively humdrum, run-of-the-mill cases... Nor was interviewing often protracted: 88.5 per cent of suspects were questioned on only one occasion and almost three-quarters of all interviews were concluded within half an hour. Although the range is, as one would expect, very wide (in this sample, from one minute to almost seven hours)."

4. In 2007, Kassin et al.²⁶ surveyed 631 police investigators on their interrogation beliefs and practices, from 16 police departments in five American states (N=574) and customs officials from two Canadian provinces (N= 57). Based on their own experience, participants estimated that the mean length of interrogation was 1.60 hours (Med =1; Range= 0 to 7; SD=.89; N= 601)... However, the range of average is represented by one standard deviation below and above the mean. Using this metric, the average length of interrogation is .71 to 2.49 hrs (just under 43 minutes to 2½ hours is their estimated average], and estimates of their longest interrogation, which lasted an average of 4.21 hours (Med=4; Range=1 to 13; SD=2.47; N=588), results in an average range of 1.74 to 6.68 hours.

Opinion

Dr. Leo uses a unique definition of interrogation duration not used by others in the field. It includes the full amount of time an individual is in custody. His early 1996 work did not use this definition, his 2004 work assembled a nonrandom sample of purported false confession cases about which very little information was known, including interrogation length, which he admits was a mixture of time in custody and actual interrogation time. Without explanation, or empirical inquiry into whether the two are the same, Dr. Leo simply adopted the definition going forward. Comparisons he then makes to other studies are therefore irrelevant, and the conclusion that the respective durations of Mr. Fulton's and Mr. Mitchell's interrogations were "extraordinary" misplaced. The length of interrogation in these cases remains a disputed fact, but there is no doubt that time in custody is not tantamount to time in actual interrogation. Based on a review of the empirical literature, including investigations Dr. Leo's cites, his claim that most interrogations last under an hour is not borne out, and it is clear that the manner in which such estimates were derived is unreliable.

V. Age as a Risk Factor

Of Mr. Fulton, Dr. Leo notes (p. 45):

"In addition to the *situational* risk factors described in Mr. Fulton's account—physical coercion and abuse, extraordinary lengthy interrogation, extreme sleep deprivation, false evidence ploys, minimization, maximization, and explicit and implicit threats and promises—Mr. Fulton, who was 18 years old at the time of the interrogation on March 18 through 21, 2003, was at a heightened risk of making and/or agreeing to a false and/or unreliable confession statement because of his personality traits and characteristics (i.e., *personal* risk factors), specifically his relative youth and secondarily his Attention Deficit Disorder (now referred to by mental health professionals as Attention Hyperactivity Deficit Disorder²⁷)."

²⁶ Kassin, S, Leo R, Meissner C, Richman K, Colwell L, Leach A, et al. (2007). Police interviewing and interrogation: A Self-report survey of police practices and beliefs. *Law and Human Behavior*, 31, 381–400.

²⁷ The DSM-5-TR Diagnosis is Attention Deficit/Hyperactivity Disorder.

1. Gross et al. (2005) reviewed 340 exoneree cases between 1989 and 2003, involving 51 false confessions.

Among adults, age 18 and older, 37²⁸ of 307 cases were identified as involving false confessions. This represents 11% of the entire exoneree sample, 73% of all false confession cases, and 13% of cases in the 18 and older age group.

Among all juveniles, ages 17 and younger, 14 of 33 cases (42%) were identified as involving false confessions. This represents 4% of the entire exoneree sample, 27% of all false confession cases, and 42% of all juvenile false confession cases.

Among juveniles 12-15, 9 of 13 cases were identified as involving false confessions. This represents 3% of the entire exoneree sample; 27% of the entire juvenile exoneree sample; 18% of all false confession cases; 64% of all juvenile false confession cases; and 69% of false confession cases in the 12-15 year age group.

Among juveniles 16-17, 5 of 20 cases were identified as involving false confessions. This represents 1% of the entire exoneree sample; 15% of the entire juvenile exoneree sample; 10% of all false confession cases; 15% of all juvenile false confession cases; and 25% of false confession cases in the 16-17 year age group.

2. Of the 375 DNA-exonerated cases listed by The Innocence Project as of July 2020, false confession was considered a factor contributing to wrongful conviction in 29% of cases (case numbers not provided; 108 or 109 cases).

Among those age 19 and older, 75 cases were identified as involving false confessions. This represents 20% of the entire exoneree sample, 69% of all false confession cases. Total case numbers for the 19 and older age group are not provided.

Among those age 18 or younger, 33 or 34²⁹ cases were identified as involving false confessions. This represents 9% of the entire exoneree sample, 31% of all false confession cases. Total case numbers for the 19 and older age group are not provided.

Data are also provided for those 21 or younger; 49% (53 cases) age 21 or younger were identified as involving false confessions, which indicates 19 or 20 cases were ages 19-21, representing 17-19% of all cases involving false confessions.

3. Of the 2991 cases listed by The National Registry of Exonerations as of February 2022, age-related false confession data are only available for 2400 cases through 3/17/20. False confession was considered a factor contributing to wrongful conviction in 292 (12%) of the 2400 cases.

Among those 18 years or older, 216 of 2189 cases were identified as involving false confessions. This represents 9% of the entire exoneree sample; 74% of all false confession cases; and 10% of false confession cases in the age 18 and older group.

²⁸ Note. Gross et al. report the 13% proportion, without case numbers. With 51 false confession cases, 33 of which are juveniles 17 and younger, there should be 37 adult case. But 37 of 307 is 12%.

²⁹ Note. The data provide the 31% proportion, without case numbers.

Among juveniles 17 years or younger, 76 of 211 cases were identified as involving false confessions. This represents 3% of the entire exoneree sample; 26% of all false confession cases; and 36% of false confession cases in the age 17 and younger group.

4. Some studies have not found age to be a significant factor (Neubauer, 1974; Mitchell, 1983; Moston et al., 1992).
5. In 1998, Pearse et al.³⁰ examine differences in psychological vulnerabilities between suspects who did vs. did not confess during police questioning. Interviews of 160 suspects were conducted from December 1991 to April 1992 at two London, England police stations. Interviews were audiotaped and transcribed, from which interview tactics were then coded. Confessions were made in 58% of the cases (50% for vulnerable and 60% for non-vulnerable, which was not a significant difference).

VII. Attention Deficit/Hyperactivity Disorder (ADHD) as a risk Factor

According to Dr. Leo (p. 45):

“...Mr. Fulton was diagnosed in childhood with Attention Deficit Disorder (now referred to as Attention Hyperactivity Deficit Disorder). Several studies have demonstrated that individuals with ADHD are at greater risk for making and agreeing to interrogation-induced false confessions.”

He opines (p. 74):

“John Fulton was at a heightened risk for making and/or agreeing to involuntary and/or unreliable statements, admissions and/or confessions during his interrogation on March 18-21, 2003 due to his personality traits and characteristics (i.e., *personal* risk factors), specifically his youth, associated psychosocial immaturity at the time, and ADD (now referred to as ADHD).”

1. Gudjonsson et al. (2008) found that the rate of self-reported false confessions was significantly higher among prisoners who were currently symptomatic for attention deficit hyperactivity disorder (ADHD) than among the other prisoners (41 and 18%, respectively). These findings highlight the potential vulnerability during questioning of people who are currently symptomatic for ADHD.”

The methodology used in this investigation is unfortunately unreliable as it rests on the self-report of inmates whose claims of innocence, false confession and psychiatric symptoms must be deemed suspect.

2. In 2012, Gudjonsson et al.³¹ surveyed 22,226 young persons in Iceland with the goal of identifying certain risk factors associated with police arrest and false confessions and to assess whether severity of ADHD symptoms increased that risk. Forty nine percent of participants were drawn from the last three years of 144 high schools (representing 86% of all mandatory students in the country) and the other 51% were college students.³² Forty nine percent identified as male. Participants were stratified into two age groups, ages 14–16 and 17–24. Self-report questionnaires were used to screen for conduct disorder and ADHD, and participants were asked whether they had ever been diagnosed with ADHD and if they were currently taking ADHD medication. Severe ADHD was defined as either a) meeting the ADHD combined type criteria (inattentive/hyperactive/impulsive) or b) those

³⁰ Pearse, Gudjonsson, Clare, Rutter (1998).

³¹ Gudjonsson G, Sigurdsson J, Sigfusdottir I, Asgeirdottir, González R, Young S. (2016). A national epidemiological study investigating risk factors for police interrogation and false confession among juveniles and young persons. *Soc Psychiatry Psychiatr Epidemiol*, 51, 359-367.

³² Gudjonsson noted that in Iceland 95% of individuals go to college after completing their mandatory education.

currently on medication and still meeting ADHD criteria of any type. Offense history was defined as conduct in the past 12 months and participants rated the frequency with which they had engaged in minor theft, major theft, violence, vandalism, and burglary during that time period. Finally, participants were asked how many times they had been interrogated at a police station in connection with a criminal offense and if they had ever confessed during such an interrogation to an offense they had not committed.

Ultimately data from 21,260 participants were available for analysis. Results indicate 14% (2987 participants) reported having been interrogated at a police station, with males reporting significantly higher rates than females (19.8% of all males vs. 8.7% of all females), and those ages 17-24 reporting significantly higher rates than those ages 14-16 (21.3% versus 10.1%). Confession data were available for 2947 of the 2987 participants who reported being interrogated; 14.7% (434) reported having made a false confession in the past, with males being significantly more likely to do so (16.2% of males versus 11.4% females among the 2947 interrogated), and those ages 14-16 were significantly more likely to make a false confession than those ages 17-24 (20.0% versus 10.3% of the 2947 interrogated).

Regarding ADHD, 10.8% (2288 of the 22,226 participants) said they had historically been diagnosed with ADHD, and 4.5% (946) reported taking medication currently, though just 4.9% (1097) met screening criteria for ADHD (i.e., were symptomatic at the time of the study), with the following breakdown by type: predominantly inattentive (2.1%), predominantly hyperactive/impulsive (1.2%) and combined type (1.6%). Of the total sample of 22,226 participants, 14.3% (3098) met the screening criteria for Conduct Disorder and 20% (4207) reported having committed one or more offenses during the past 12 months (minor theft, major theft, violence, vandalism, or burglary).

Of the 14% of the sample (2987) that had been interrogated, the variables that produced the best “odds”³³ of being interrogated were a) meeting the screening criteria for Conduct Disorder (39.4% of the 2987, effect size 5.9); and b) meeting the current criteria for ADHD combined type (4.5%, effect size 4.3). An effect size of 4.0 and greater is considered strong. Regarding risk of making a false confession, none of the variables produced a strong effect size. This is notable given a sample size of the magnitude in this investigation lends itself to producing statistically significant but clinically insignificant findings/inflated effect sizes. Moderate effect sizes were found for a) currently being on medication for ADHD (24.1% of 434 false confessions, effect size 3.9); and b) currently meeting the screening criteria for ADHD combined type (11.3%, effect size 3.7).

Comment. First and foremost, it must be pointed out that this investigation did not assess if any of the interrogations in which participants engaged were for the offense of murder, the only relevant type of interrogation in Mr. Fulton’s and Mr. Mitchell’s cases given the very different stakes for most other offense types. Dr. Leo overstates the relative risk of making a false confession given age and a history of an ADHD diagnosis in any case based on these data. While those ages 14-16 undergoing interrogation were significantly more likely to make a false confession compared to those ages 17-24 (20.0%, 589 participants versus 10.3%, 304 participants), the overall rate of false confession among all interrogated participants ages 14 to 24 was 14.7%, a number that, while certainly meaningful, is relatively small, making its application to single cases like Mr. Fulton and

³³ The authors used odds ratio statistics to determine effect sizes for each variable. The odds ratio is used to determine if a particular variable is a risk factor for the outcome of interest and provides the magnitude of its risk. For example, the Odds Ratio compares the relative odds of confession vs. no confession given exposure to the variable of interest (e.g., age). An Odds Ratio of 2 is the recommended minimum effect size representing a “practically” significant effect for social science data, 3.0 is a moderate effect, and 4.0 is a strong effect.

Mr. Mitchell's difficult to understand. The same can be said for ADHD. Dr. Leo points out that Mr. Fulton historically was diagnosed with ADHD during childhood. Gudjonsson et al. did not find a history of an ADHD diagnosis to be a strong predictor of false confession, nor a moderate one (effect size 2.7).

Generalizing from Group Data to Individuals

Dr. Leo's application of group data to an individual is problematic; even if one assumes Mr. Fulton's or Mr. Mitchell's case is one of false confession; such extrapolations have great potential to be misrepresentative and misleading. In all research, and in all applied science (e.g., neuropsychology), findings come with confidence intervals that reflect the potential for measurement error, an accepted aspect of group to individual extrapolations. These error rates can be significant, such as in IQ scores. Depending on the level of confidence one wishes to employ (e.g., 90% versus a more stringent 95%, the latter being typical in the social sciences, versus 99% in medicine in order to reduce false negative rates and missing disease when it exists), the range of possible "truths" about an individual can be significant. An individual's obtained IQ score is therefore always represented within this range of possible "truths," e.g., an IQ of 100 will fall somewhere between the range of say 96 – 104 at a 90% confidence interval, versus 93 – 107 at a 95% confidence interval, etc. Because the error rates in false confession research are not known, attempting to find Mr. Fulton's place within a group of say, others with similar characteristics in false confession samples (or with similar false confession "risk factors") is not possible.

Sample Size for Meaningful Group Comparison

Social scientists such as Dr. Leo have drawn critical inferences from the data collected about false confessions based on woefully small data points. The field research has clear directional hypotheses (e.g., youth, intellectual disability or mental illness are risk factors for false confessions), but there are simply too few data points to conduct any meaningful analyses within the sample. The adult sample of DNA-exonerated cases is currently approaching an appropriate number of false confession cases for *nonparametric* tests (say to compare frequency data/its false confession numbers with those of another group), but no comparison group is available as they are all far too small (e.g., a comparison between the number of juveniles versus adults making false confessions cannot be conducted at this time). And the sample sizes needed to conduct *parametric* tests (say to compare group means on some variable between adult versus juvenile groups) are far greater, approximately 100-200 for a two-group comparison, and up to 500 or more per group for three-group comparisons if post hoc tests are conducted (say one wanted to compare adults versus juveniles age 16-17 versus juveniles age 12-15 on some variable).

Clearly this research is in its nascent form, and we will learn a great deal from the datasets as they grow. But at this time, there are no compelling data, as intuitive as it may seem, to conclude with any statistical certainty that juveniles make false confessions at significantly greater rates than adults, or that any of the other identified "risk factors" do in fact significantly increase confession risk from a statistical standpoint. While trends in data are interesting, as in all research, hypotheses about their meaning must be tested, and ultimately replicated, to ensure reliability and validity. The conclusions Dr. Leo has reached based on the available data must be considered unreliable, and the "risk factors" he identifies very preliminary. His methodology has no place in an applied setting, in my opinion, and has the great potential to be misleading and confusing in a courtroom.